Are inflation forecasts from major Swedish forecasters biased?*

Michael Lundholm
Department of Economics, Stockholm University†

June 3, 2010

Abstract

Inflation forecasts made 1999–2005 by Sveriges Riksbank and Konjunkturinstitetet of Swedish inflation rates 1999–2007 are tested for unbiasedness; i.e., are the mean forecast errors zero? The bias is in the order of -0.1 percentage units for horizons below one year and in the order of 0.1 and 0.6 (depending on inflation measure) above one year. Using the *maximum entropy bootstrap* for inference bias is significant whereas inference using HAC indicates insignificance.

JEL: E37.

Keywords: Forecast evaluation, inflation, unbiasedness, maximum entropy bootstrap.

1 Introduction

That forecasts are unbiased, that is, there is no systematic over- or underprediction, is often seen as a desirable property (Clements, 2005, p. 4). As is well known unbiasedness is a necessary condition for optimal forecasts when the forecaster minimise a quadratic error loss function; i.e., when forecasts are chosen so as to minimise the mean square forecast error (MSFE). Given such a loss function, rejecting the null hypothesis of unbiasedness implies rejecting that the forecasts are optimal.

Is unbiasedness what we should expect for the inflation rate forecasts from the two leading official Swedish forecasters of inflation, Sveriges Riksbank (the Central Bank of Sweden) (RB) and Konjunkturinstitutet (the

*I am grateful for comments from Ulf Holmberg and other seminar participants at Umeå University.

†Address: Department of Economics, Stockholm University, SE 106 91 Stockholm, Sweden. Email: michael.lundholm@ne.su.se.

National Institute of Economic Research) (KI)?[1] Of particular interest are the forecasts of RB which employs a forecast based monetary policy. The RB has as its objective a symmetric inflation target of a 2 percent increase in consumer prices with a tolerance band of ± 1. To expect unbiased forecasts from the RB therefore seem to have some *ex ante* support. Testing for unbiasedness in these forecasts therefore seems motivated. In addition, a characterisation of the possible bias in these forecasts may be very useful for users of these forecasts (other government agencies, households and firms) in their economic decisions.

However, the RB employs a judgemental forecasting procedure where the probability distribution of the point forecasts is a (potentially) asymmetric two–piece normal distribution.[2] Therefore, the RB does not explicitly forecast by minimising the MSFE and if they had done so is unbiasedness not necessarily an indication of optimality. Also KI applies judgemental forecasting but this procedure is not as well documented and the *ex ante* reasons for and against unbiased forecasts are not as obvious as in the case of RB.

Previously Jansson and Vredin (2003), although not explicitly testing for unbiasedness, found that inflation forecasts from RB for the period 1992–1998 overpredicted the inflation rate. Although the actual inflation rate decreased from 5 to less than 1 percent during these years, they attributed this overprediction to the conditioning assumption of the RB (that their main policy rate is unchanged) and/or the judgemental forecasting procedure.

In the present study tests for unbiasedness are carried out on RB forecasts for the inflation rate 1999–2007 (see Figure 1 on page 9 for the inflation outcome) and on KI forecasts for the inflation rate 2001–2007, measured as relative 12 month change in CPI (the standard consumer price index) and KPIX (CPI with temporary effects excluded). These forecasts were made quarterly by RB 1999–2005, with forecast horizons $1 - 25$ months, and KI 2001–2005, with forecast horizons $1 - 21$ months; in the case of RB conditional on an unchanged policy rate.

Since the forecasts exhibit typical time series data properties, inference using the OLS estimate of the sampling variance of the mean forecast error is not valid. To take these problems into account inference is made using hetereoskedasticity and autocorrelation consistent standard errors (HAC) and maximum entropy (ME) bootrapping.

Focusing at statistical significance the main result is that using the ME bootstrap both RB and KI produce biased forecasts over almost all forecast horizons. There is a systematic tendency for both forecasters to underpredict the inflation up to one year ahead and overpredict it about one to two years ahead. Inference from the HAC estimates typically lead to statistical

[1] Both institutions are independent Swedish government authorities, RB directly under the parliament and KI under the government.

[2] See Blix and Sellin (1998, 1999) and Berg (2000) for a documentation of the RB forecasting procedure.

insignificant results.

The magnitudes of the mean of forecast errors are in the order of -0.1 percentage units for forecast horizons below 12 months (see Figure 2 on page 10) and 0.6 percentage units above 12 months for CPI and 0.2 percentage units for KPIX (see Figure 3 on page 11) for both RB and KI. Using results from Jansson and Vredin (2003) one can conclude that a positive bias of about 0.6 percentage units transforms into a 0.5 to 0.9 percentage points higher policy rate, compared to unbiased forecasts. This is not the only consequence for households and firms. Since they also may act on the forecast the bias may also have other indirect economic consequences, depending on their loss functions when using the forecasts. Therefore, the estimated bias seems to be economically important.

This paper is organised as follows: Section 2 describes the data set and the testing strategy. The different methods for inference are compared in section 3 where the main results are presented. The paper is concluded in section 4.

All estimations were performed using R version 2.11.0 (2010-04-22).[3] Lundholm (2010a) is a technical documention accompanying this paper. It contains the econometric code with comments, detailed information about versions of the econometric software and packages and a more detailed presentation of the results.

2 Data and testing strategy

Data (forecasts and corresponding outcomes) are publicly available as the R package **sifds**.[4] The package and its availability as well as the data and its sources are described in greater detail in Lundholm (2010c).

In package **sifds** data consists of yearly inflation rate forecasts from the two forecasters RB and KI. Inflation rates are measured as relative 12 month change in CPI (the standard consumer price index) as well as KPIX (CPI with temporary effects excluded) for each month during the forecasted time periods. These time periods differ between forecasters; the period 1999:M05–2005:M07 is covered by RB and 2001:M04–2005:M07 by KI; see Figure 1 for the inflation rates during the forecasted period. The forecasts were published with a 3 month interval between the origins. For RB the forecast origins span the period 1999:Q2–2005:Q2 and for KI 2001:Q1–2005:Q2.[5] The forecast horizons also differed between forecasters. RB had

[3] See R Development Core Team (2011).

[4] Lundholm (2010b).

[5] The forecasters have continued to produce forecasts after 2005:Q2, but the data set does not include these because RB changed (i) a basic assumption underlaying the forecasts and (ii) the number of forecast origins (i.e. the frequency) from 4 to 3 each year. The basic assumption that changed in 2005 was that rather assuming that their main policy rate would remain unchanged over the forecast horizon the RB started to forecast its own

horizons from 1 month up to 25 or 26 months and KI horizons from 1 month up to between 21 and 30 months.

This means that with 25 forecast origins for RB and 18 for KI there are at a maximum 25 or 18 observations in each time series of forecasts with the same horizon from the same forecaster; for the longest horizons considerably less. In order to have no less than 25 observations in the RB forecast series and 18 in the KI forecast series, all RB forecasts for the 26 months horizon and all KI forecasts with horizons 22 months and more are removed from the data set in **sifds**.

The forecast errors are as usual defined as

$$\mathsf{FE}_{t+h|t} = \mathsf{F}_{t+h|t} - \mathsf{O}_{t+h}, \tag{1}$$

where $\mathsf{F}_{t+h|t}$ is the h–month ahead forecast made at time t, O_{t+h} is the inflation outcome for time $t+h$ and $\mathsf{FE}_{t+h|t}$ is the corresponding forecast error. The RB data consists (for CPI as well as KPIX) then consist of 25 such vectors of forecast errors (one for each horizon $h \in \{1, 2, \ldots, 24, 25\}$) with 25 observations in each and the KI data consists (also for CPI and KPIX) of 21 vectors of forecast errors (one for each horizon $h \in \{1, 2, \ldots, 20, 21\}$) with 18 observations in each.

Since unbiasedness is the same as a zero mean forecast error a formal test procedure could be to estimate

$$\mathsf{FE}_{t+h|t} = \mathsf{MFE} + e_{t+h}, \tag{2}$$

with OLS, where the intercept MFE is the the mean forecast error and e_{t+h} are the estimated residuals, and then test the null hypothesis that $\mathsf{MFE} = 0$ against the alternative that it is not.[6]

There are $2 \times (25 + 21) = 92$ different forecast series to be tested. Summary statistics are given as box–plots in Figures 2–3 on pages 10–11.[7] From the graphs wee see that means taken over the MFE's are negative but close to zero (about -0.1)for forecast horizons up to 12 months (Figure 2) for both forecasters and both inflation measures. For forecast horizons above one year (Figure 3) the mean of the MFE's are positive and large in the case of CPI (larger than 0.5). If we study the distribution of all 92 mean forecast errors (Figure 4 on page 12) wee see that we have the highest frequencies

policy rate.

[6]This procedure is identical to estimating $\mathsf{F}_{t+h|t} = \mathsf{MFE} + \beta\mathsf{O}_{t+h|t} + e_{t+h}$ and test the simultaneous null hypothesis $\mathsf{MFE} = 1 - \beta = 0$; see Holden and Peel (1990).

[7]The box–plots (or "box–and–whiskers"–plot) show the mean as a black line within a box, the length of which equals the inter–quartile range, with "whiskers" extending to the most extreme MFE which is no more than 1.5 times the inter–quartile range away from either the first or third quartiles. Had there been any MFE more extreme than those shown by the "whiskers" had they been shown in the graphs. The "whiskers", therefore, here show the minimum and maximum MFE's.

in the interval $[-0.2, 0.1]$ and that the top tail above 0.1 tend to be thicker than the bottom tail below -0.2.

Testing the null hypothesis that the MFE's are drawn from a normal distribution also gives the p–value $3.242e-05$.

For OLS to be used for inference the regression residuals have to satisfy the standard assumptions about homoskedasticity an no serial correlation. Since the residuals are affine transformations of the forecast errors, this requirement also applies to the forecast errors. However, it turns out that most of the forecast error series can be used to estimate a standard ARIMA model. Table 1 shows the number of forecast error series (divided on categories forecaster and inflation measure) for which the model ARIMA$(0, 0, 0)(0, 0, 0)$ is correct. Only 12 series, out of the 92 series, have this property and only one (1) series if we just consider forecast horizons above 12 months. Instead, this is a clear indication that OLS is unsuitable for inference in this case. See Lundholm (2010a) for details about testing and identification procedures.

Since only OLS would be able to use on a small part of the data alternative methods for inference have to be used. Here we use (i) hetereoskedasticity and autocorrelation consistent standard errors (HAC) and (ii) maximum entropy (ME) bootrap.[8] The OLS results are, however, also reported for comparison.

3 Results

The quantiles for 2.5% and 97.5% around the MFE are calculated for RB and KI for both inflation measures CPI and KPIX and all three inference techniques (HAC, ME bootstrap and OLS) for each forecast horizon. The results are presented in the graphs in Figure 5 on page 13, where the MFE is a white line and the 95% confidence intervals according to HAC, ME bootstrap and OLS are depicted in different shadings of grey. The numerical values of the quantiles are reported in Lundholm (2010a).

The difference inference methods give distinctly different results:

- Using HAC standard errors the conclusion is that the null hypothesis of unbiasedness cannot be rejected in most cases. Exceptions when

[8]For HAC covariance matrix estimators see Newey and West (1987, 1994) and for their implemention in R see Zeileis (2004).

For ME bootstrap see Vinod (2006); Vinod and de Lacalle (2009). That some series are non–stationary makes e.g. block bootstrap impossible to use. The ME bootstrap algorithm does not, however, require stationarity. It allows draws (i) close to original values (which is reasonable with time structured data), (ii) outside the closed interval defined by the sample minimum and maximum and (iii) preserves the time structure of data. basically data is reordered in increasing values. Around each observation is then an interval formed bounded by the nearest lower and higher values and it is from this interval the bootstrapped value replacing the original observation is drawn. The mainlimitation of ME bootstrap, that it cannot create meaningful ensembles for binary variables, does not apply here.

the null is is rejected are the forecasts RB CPI $h \in \{25\}$, KI CPI $h \in \{1, 18, 21\}$ and KPIX $h \in \{1\}$.

- The ME bootstrap gives different results: Up to a one year forecast horizon forecasts are basically negatively biased and above one year positively biased. Forecaster or inflation measure do not matter. Exceptions are the forecasts RB CPI $h \in \{8, 9\}$, KPIX $h \in \{10, 13\}$, KI CPI $h \in \{4, 6, 7, 10, 11\}$ and KPIX $h \in \{5, 13\}$ for which the null cannot be rejected. These are the intermediate horizons where (basically) shorter horizons have negative bias and longer horizons positive bias.

- Had the OLS results been used for inference the main conclusion would again have been different: For CPI forecasts from the two forecasters up to an horizon of about one year (12 months) the null would have been rejected and for those above one year the null would *not* have been rejected with significant positive bias. For KPIX the conclusion would have been even more clearcut for not rejecting the null hypothesis of unbiasedness, except for a handful of forecasts with the longest horizons.

4 Discussion

The main results are that ME bootstrapping shows that the inflation forecasts evaluated are significantly biased. Negatively biased up to a one year forecast horizon and positively biased above one year. On the other hand, inference using HAC did not show any significant bias, except in a handful of cases. One explanation for these diverging results may be the autocorrelation in the forecast errors. A positive autocorrelation implies that the distance between sequential observations tend to be small. In ME bootstrap the interval, from which replicas of a certain observation are drawn, will then tend to be small which makes the resulting MFE replicates similar to the original and the resulting confidence band narrow. In HAC a (positive) autocorrelation will add on to the estimated sampling variance and create wider confidence bands. However, the ME bootstrap results are consistent with what have previously been reported by Jansson and Vredin (2003).

What are the consequences of such as bias? Since it takes $1 - 2$ year before a change in the policy rate has full impact on the inflation rate only the longer forecast horizons are of interest. Considering simple policy rules of how changes in policy rates are affected by (among other things) changes in the forecasted inflation rate we may get an indication how different the monetary policy had been had forecasts been unbiased. A forward–looking theoretical Taylor–type rule (Rudebusch and Svensson) when the loss function for policy choice puts equal weight on inflation and unemployment, transforms a unit change in forecasted inflation rate into 1.5 unit change in the policy rate. Estimating the same model using data 1992-1998 Jansson and Vredin (2003) found that a unit change in the forecasted inflation increases the policy rate with 0.81 units. Unbiased forecasts would have been on average about 0.6 (or more) percentage points lower then the actual and therefore the policy rate would have been at least about between 0.5 to 0.9 percentage points lower than the actual.

Which are the possible explanations for these results? One explanation is that both forecasters are (implicitly using) asymmetric loss functions when constructing the forecasts. Both forecasters apply judgemental forecasting and in the procedure of transforming the more model based forecasts (where such an asymmetry is unlikely unless explicitly modelled) to the published forecasts such a bias may enter. Another explanation is the actual time period investigated; see Figure 1 on page 9. In 2001 there was a drastic increasing in the inflation rate for about two years when inflation even more drastically dropped to negative numbers in 2004 and 2005. This implies that many forecast origins during 2001–2003 occurred during relatively high inflation rates whereas the the longer forecasts horizons finished during a low inflation rate period. Which is a similar pattern as during the period 1992-1998 examined by Jansson and Vredin (2003).

References

Berg C (2000). "Inflation forecast targeting: The Swedish experience." In MI Blejer, A Ize, AM Leone, S Werlang (eds.), "Inflation targeting in practice: Strategic and operational issues and application to emerging market economies," pp. 28–36. International Monetary Fund. URL `www.imf.org/External/Pubs/FT/seminar/2000/Targets/strach4.pdf`.

Blix M, Sellin P (1998). "Uncertainty bands for inflation forecasts." Working paper 65, Sveriges Riksbank, URL `http://www.riksbank.se/upload/1013/98nr65.pdf`.

Blix M, Sellin P (1999). "Inflation forecasts with uncertainty intervals." *Penning– och valutapolitik*, **1999**(2), 12–28. URL `http://www.riksbank.se/upload/Dokument_riksbank/Kat_publicerat/Artiklar_PV/qr99_2_artikel2.pdf`.

Clements MP (2005). *Evaluating econometric forecasts of economic and financial variables*. Palgrave.

Holden K, Peel DA (1990). "On testing for unbiasedness and efficiency of forecasts." *The Manchester School*, **58**, 120–127. URL `http://dx.doi.org/10.1111/j.1467-9957.1990.tb00413.x`.

Jansson P, Vredin A (2003). "Forecast–based monetary policy: The case of Sweden." *International Finance*, **6:3**, 349–380. URL http://dx.doi.org/10.1111/j.1367-0271.2003.00122.x.

Lundholm M (2010a). "Are inflation forecasts from major Swedish forecasters biased? Technical documentation." Unpublished manuscript, Department of Economics, Stockholm University, URL http://people.su.se/~lundh/projects/unbiased/unbiased_techreport.pdf.

Lundholm M (2010b). *sifds: Swedish inflation forecast data set.* R package version 0.9, URL http://www.cran.r-project.org/web/packages/sifds/index.html.

Lundholm M (2010c). "sifds: Swedish inflation forecast data set 1999:Q2–2005:Q2." Research paper in economics 2010:4, Department of Economics, Stockholm University, URL http://swopec.hhs.se/sunrpe/abs/sunrpe2010_0004.htm.

Newey WK, West KD (1987). "A Simple, Positive semi-definite, hetereoskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, **55**, 703–708. URL http://www.jstor.org/stable/1913610.

Newey WK, West KD (1994). "Automatic lag selection in covariance matrix estimation." *Review of Economic Studies*, **61**, 631–653. URL http://www.jstor.org/stable/2297912.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Vinod HD (2006). "Maximum Entropy Ensembles for Time Series Inference in Economics." *Journal of Asian Economics*, **17**, 955 – 978. URL http://dx.doi.org/10.1016/j.asieco.2006.09.001.

Vinod HD, de Lacalle JL (2009). "Maximum Entropy Bootstrap for Time Series: The meboot R Package." *Journal of Statistical Software*, **29**(5). URL http://www.jstatsoft.org/v29/i05/paper.

Zeileis A (2004). "Econometric computing with HC and HAC covariance matrix estimators." *Journal of Statistical Software*, **11**(10). URL http://www.jstatsoft.org/v11/i10/paper.

Figure 1: Yearly changes in consumer prices (CPI and KPIX) 1999:M5-2007:M07.

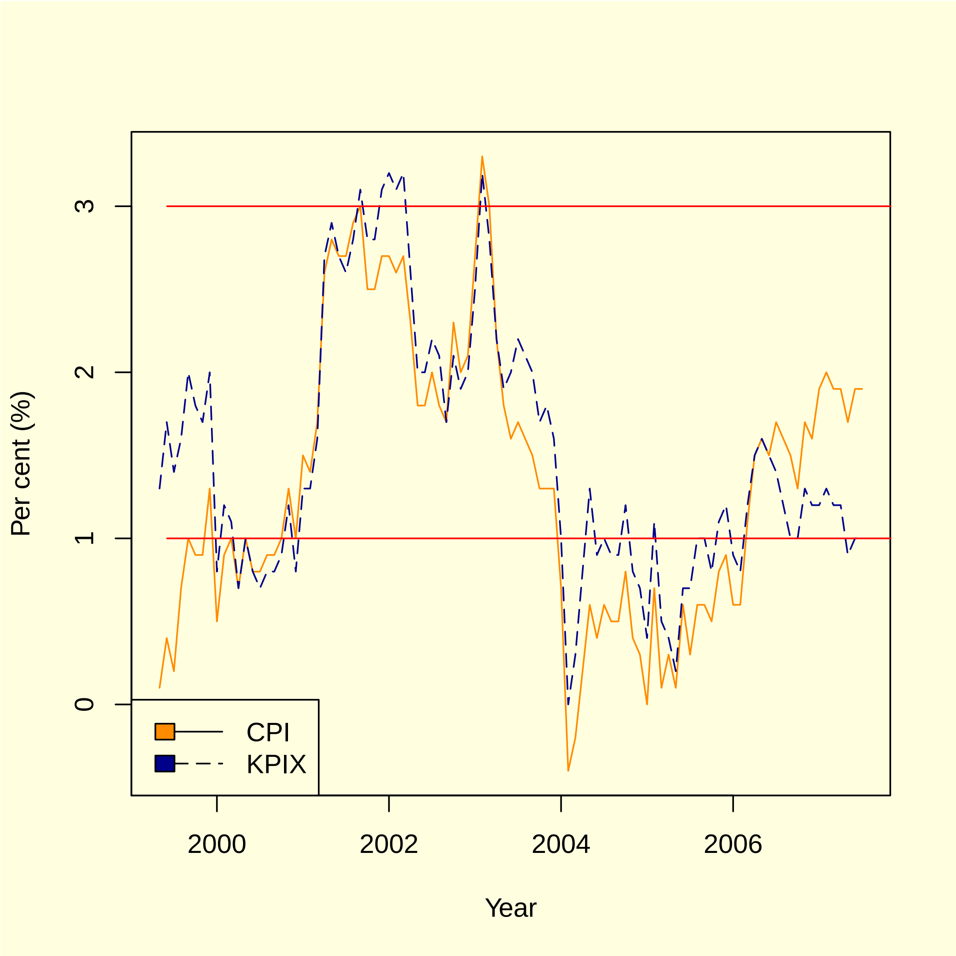

Table 1: Number of MFE series for which ARIMA(0, 0, 0)(0, 0, 0) is the correct model

	$h \leq 12$	$h > 12$
RB CPI	2	0
RB KPIX	1	0
KI CPI	4	0
KI KPIX	4	1

Figure 2: Boxplots for mean forecast errors (MFE) with horizons $h \leq 12$.

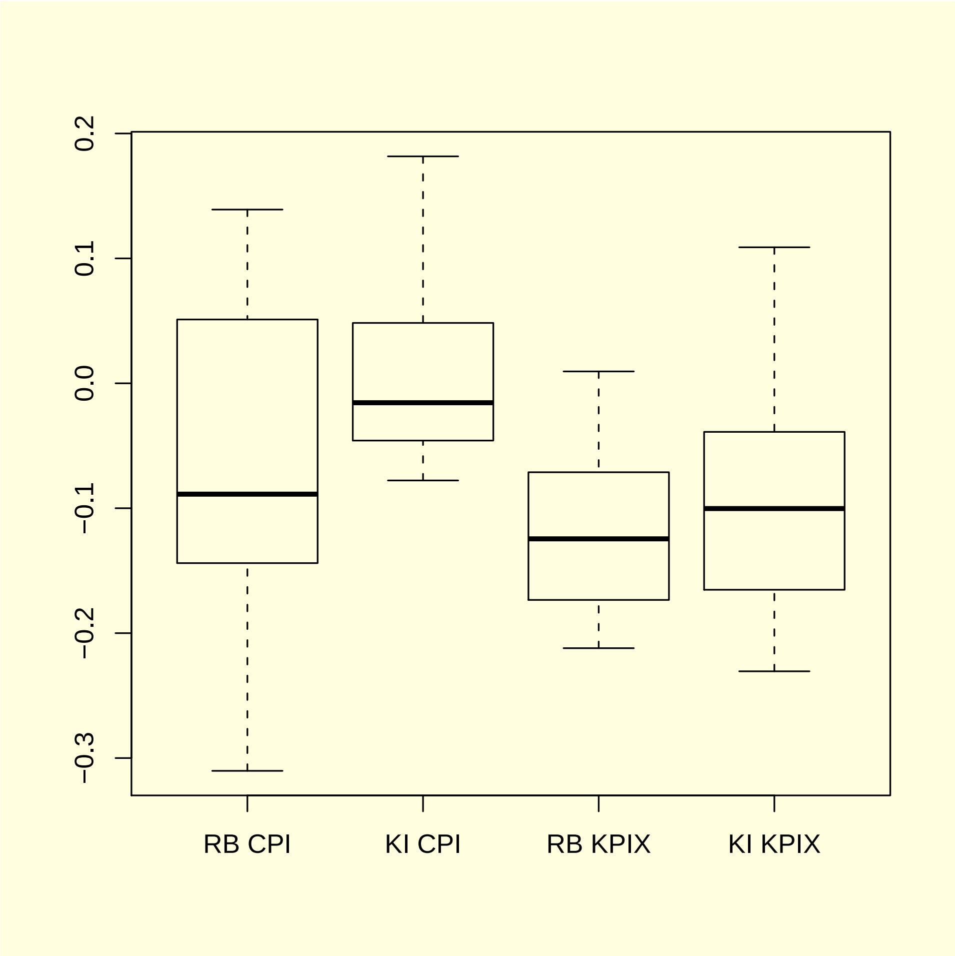

Figure 3: Boxplots for mean forecast errors (MFE) with horizons $h > 12$.

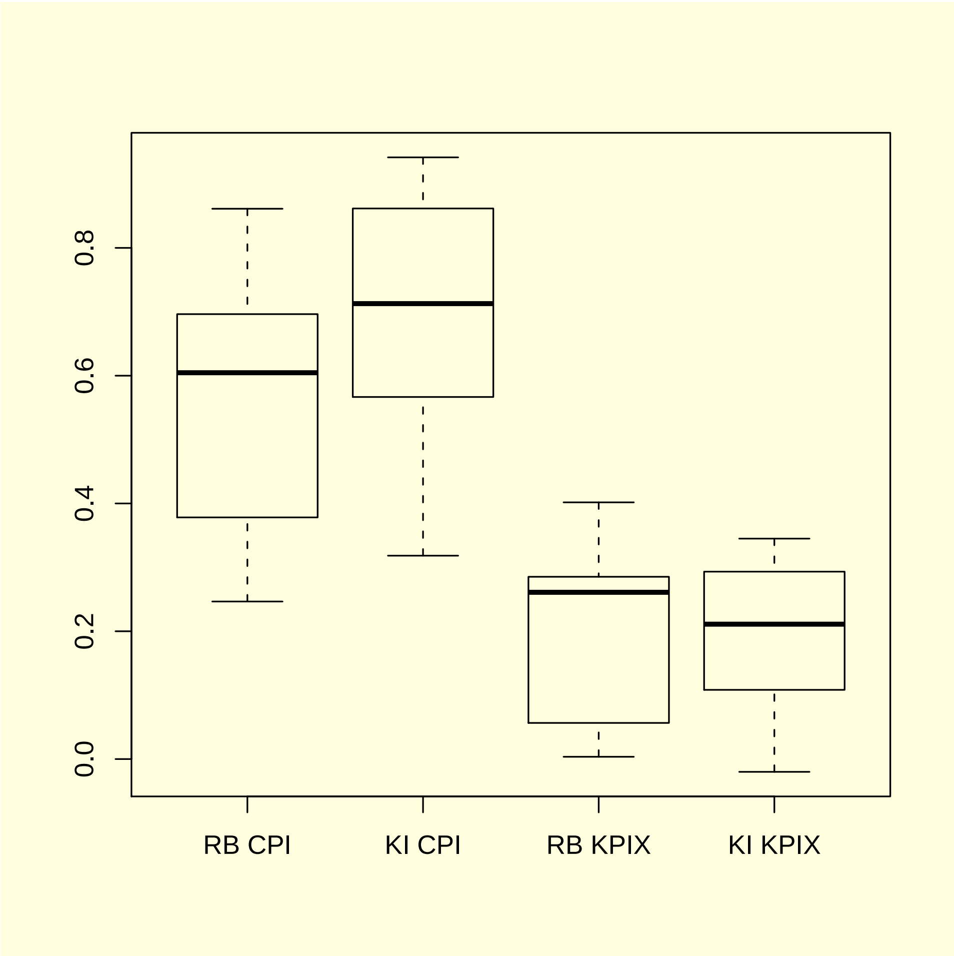

Figure 4: Histogram for all mean forecast errors (MFE) $N = 92$.

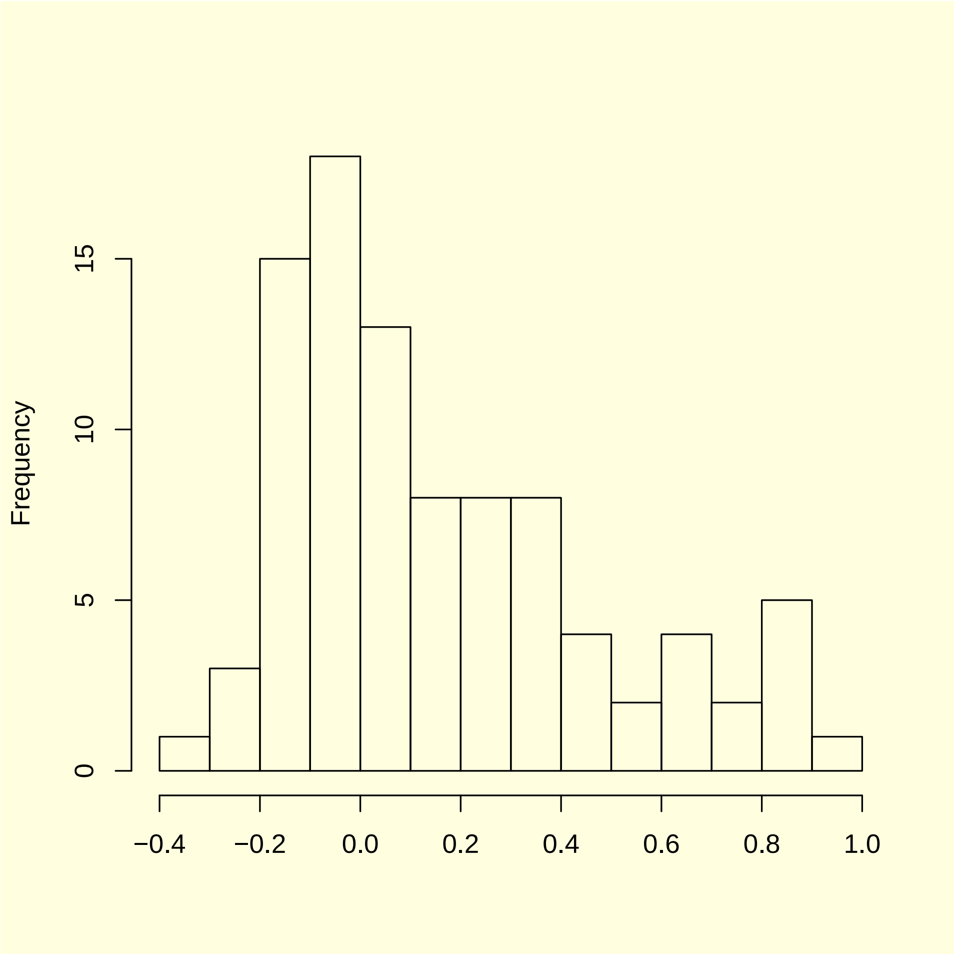

Figure 5: Mean forecast errors (MFE) over all horizons with 95% confidence bands.